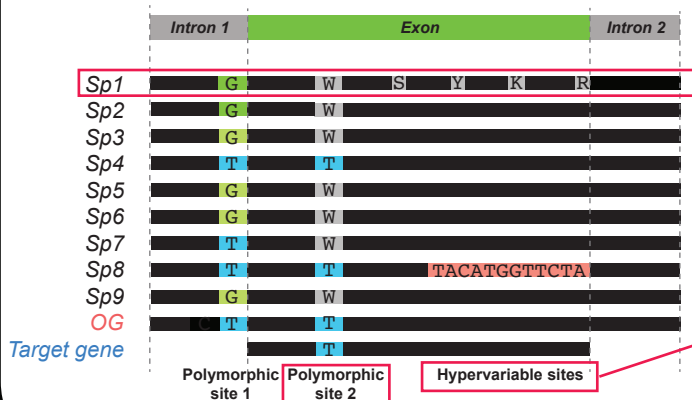


# PPD pipeline

**Part1: generate “degenerated” sequences**

**Part2: trimming and paralog detection**

## Gene: “degenerated” MSA



s1: resort gene files

s2: filter sequences with high percentage of heterozygous sites (default 0.05%).

s3: add target references to gene files.

s4: obtain MSA using MAFFT.

s5: remove the target references.

s6: remove the sites missing in >50% samples (gaps) using Trimal.

s7: remove hypervariable sites using sliding window method.

s8: identify putative paralogs using shared heterozygosity in over 50% taxa.

PPD Trimming Steps

**Paralogs Detection**